population
onomastic databases

Gerrit Bloothooft
Universiteit Utrecht / Meertens Instituut KNAW / IISG KNAW
names!

• why are they interesting

• everyone has a name
• a name is part of ones’ identity
• most people name children (to their own liking)
• names link to language, culture, social mechanisms, region: they can reveal aspects of these
• naming is dynamic with issues of time and space
• many pattern recognition challenges
topics

• huge name databases
• preparing databases
  • on-line presentation
  • availability | privacy issues
• possibilities for research
  • onomastics | social sciences | genetics
• extentions
  • infrastructure for external contributions
    – name origin, variants, etymology, spread
name databases

• available from civil registration
  • Meertens Institute KNAW & Utrecht University
  • for scientific research only
  • no identification of individuals allowed

• first names
• family names

• more than just names!
information

- **first names**
  - all first names
  - date, place and country of birth
  - current residence (postal code)
  - id
  - id of parents

- **family names**
  - prefix and family name
  - date, place and country of birth
  - current residence
full population

• first names
  • 22 million persons
    – 16 million in 2006

  [civil registration digitized in 1994]
  – 1,5 million died 1994-2006
  – 4,5 million died before 1994
    » no own record, but mentioned as parent
    » reconstruction needed because of multiple and inconsistent presence as parent

• family names
  • 16 million persons in 2007
different names

• first names
  • 500.000
  • 300.000 in first position
  • 5.000.000 as full name

• family names
  • 315.000
online name corpora

• first names
  – [www.meertens.knaw.nl/nvb](http://www.meertens.knaw.nl/nvb)
  – June 3, 2010

• family names
  – [www.meertens.knaw.nl/nfb](http://www.meertens.knaw.nl/nfb)
  – December 3, 2009
name search

- exact  
  - pick one from list
- begins with 
  - pick one from list
- ends with 
  - pick one from list
- contains 
  - pick one from list
- advanced 
  - regular expression
  - aggregated data

matthijs
matt
ijs
th
\^ma.*(ij|y)s\$
on show

all presentations absolute & relative

• first names
  • name & gender
  • # as first name, # as following name (totals 2006)
  • births per year (since 1880)
  • places of birth (for 2006 population)
    – municipalities in 2006
  • explanations (20.000, ~4000 extensive)

• family names
  • name
  • # in 2007, # in 1947
  • places of residence in 2006, provinces in 1947
  • explanations (90.000, ~4000 extensive)
  • relational network of names
Corpus of First Names in The Netherlands

The Corpus of First Names in The Netherlands presents 500,000 names. The names originate from Civil Registration (2006) enhanced with recent names that were made available by the Social Security Bank. Also included are first names from the latest edition of Prisma Voornamen (edited by J. van de Schaar, revised by Doreen Gerritzen) which are not used anymore in The Netherlands.

For many names the annual popularity distribution since 1880, and a geographic distribution is presented. For about 20,000 names an explanation can be found. The information is available in Dutch only.

Doreen Gerritzen, editor name etymology
Gerrit Bloothof (Universiteit Utrecht), GBA co-ordinator
Matthijs Brouwer, programming
Jan Pieter Kunst, application components

contact: nvb@meertens.knaw.nl
Nederlandse Voornamenbank

Gerrit

zie ook Gérrit en Ėrrit

populariteit  verspreiding  verklaring

<table>
<thead>
<tr>
<th></th>
<th>NL totaal (2006)</th>
<th>%</th>
<th>[populariteit]</th>
<th>[% populariteit]</th>
</tr>
</thead>
<tbody>
<tr>
<td>m</td>
<td>als eerste naam:</td>
<td>95900</td>
<td>1.1345%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>als volgnaam:</td>
<td>32831</td>
<td>0.4211%</td>
<td></td>
</tr>
<tr>
<td>v</td>
<td>als eerste naam:</td>
<td>&lt; 5</td>
<td>&lt; 0.0001%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>als volgnaam:</td>
<td>55</td>
<td>0.0007%</td>
<td></td>
</tr>
</tbody>
</table>

Populariteit van 'Gerrit' als eerste naam voor mannen tussen 1880 en 2009
Nederlandse Voornamenbank

Gerrit

zie ook Gérrit en Ėrrit

<table>
<thead>
<tr>
<th>Vor</th>
<th>populariteit</th>
<th>verspreiding</th>
<th>verklaring</th>
</tr>
</thead>
<tbody>
<tr>
<td>m</td>
<td>NL totaal (2006)</td>
<td>%</td>
<td></td>
</tr>
<tr>
<td>als eerste naam:</td>
<td>95900</td>
<td>1.1345%</td>
<td>[populariteit]</td>
</tr>
<tr>
<td>als volgnaam:</td>
<td>32831</td>
<td>0.4211%</td>
<td>[populariteit]</td>
</tr>
<tr>
<td>v</td>
<td>als eerste naam:</td>
<td>&lt; 5</td>
<td>&lt; 0.0001%</td>
</tr>
<tr>
<td>als volgnaam:</td>
<td>55</td>
<td>0.0007%</td>
<td>[populariteit]</td>
</tr>
</tbody>
</table>

Populariteit van 'Gerrit' als eerste naam voor mannen tussen 1880 en 2009 (percentages)
Aantal mannen met 'Gerrit' als eerste naam per geboortegemeente in 2006

1 - 4
5 - 10
11 - 25
26 - 50
51 - 100
101 - 250
251 - 500
501 - 1000
1001 - 2500
> 2500
Percentage mannen met 'Gerrit' als eerste naam per geboortegemeente in 2006
Nederlandse Voornamenbank

Gerard

zie ook Gérard, Gèrard, Gerárd, Gérard, Gèrard, Geràrd, gerard, Geràrd, Gerard en Gèrard

Verklaring

Tweestammige Germ. naam met de betekenis 'sterk met de speer', uit Ger- 'speer' (zie -ger-) en -hard 'hard, sterk, stevig' (zie -hard-).
Nederlandse Voornamenbank

**pke$**

<table>
<thead>
<tr>
<th>Voornaam</th>
<th>Mannen</th>
<th>Vrouwen</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aepke</td>
<td></td>
<td>&lt; 5</td>
</tr>
<tr>
<td>Ale-Wopke</td>
<td>&lt; 5</td>
<td></td>
</tr>
<tr>
<td>Anjakoopke</td>
<td>&lt; 5</td>
<td></td>
</tr>
<tr>
<td>Apke</td>
<td>&lt; 5</td>
<td>&lt; 5</td>
</tr>
<tr>
<td>Bapke</td>
<td>-</td>
<td>34</td>
</tr>
<tr>
<td>Bepke</td>
<td>&lt; 5</td>
<td>5</td>
</tr>
<tr>
<td>Beroke</td>
<td>&lt; 5</td>
<td>&lt; 5</td>
</tr>
<tr>
<td>Ciepke</td>
<td>&lt; 5</td>
<td>&lt; 5</td>
</tr>
<tr>
<td>Cipke</td>
<td>&lt; 5</td>
<td>-</td>
</tr>
<tr>
<td>Csepke</td>
<td>&lt; 5</td>
<td>-</td>
</tr>
</tbody>
</table>

**Geavanceerd zoeken naar namen met 'pke$'**

1 t/m 10 van 153 resultaten

Volgende 10
Percentage mannen voor reguliere expressie 'pkæ$' als eerste naam per geboortegemeente in 2006

- < 0,01%
- 0,01% - 0,02%
- 0,02% - 0,05%
- 0,05% - 0,1%
- 0,1% - 0,2%
- 0,2% - 0,5%
- 0,5% - 1%
- 1% - 2%
- 2% - 5%
- > 5%
data exploration

• some examples
  • Zipf’s law applies on names
  • name revolution
    – naming after relatives
    – naming as fashion
  • from many names, to a few groups
    – 500.000 first names are untractable
    – many names “behave” the same way
    – from 500.000 to 10-30 name groups
    – relations with social groups (and fashion)
Zipf’s law applies on names

170,000 first names of 3.5 million children (1983-2001)
from tradition to fashion

Jan

Kevin

Nelly

Jayden

Ingrid

Femke
from many to few

• assumption
  • parents chose names for their children that fit their social environment
    – traditional names (traditional Dutch | latinized)
    – Frisian, Arabic, Turkish names
    – English, French, Scandinavian, Spanish, Italian names
    – names from the Old Testament
    – names from history and culture
    – names from nature
    – ..
  • analysis of names of children found in the same family
map of first names
family names
The Corpus of Family Names in the Netherlands presents the 314,000 family names of all persons with the Dutch nationality, who lived in the Netherlands in the year 2007. Also included are about 100,000 family names registered at the census from 1947. Besides frequencies and geographic distribution maps, additional information about the meaning, composition and history can be found for many names. The names are hierarchically interconnected on the basis of spelling, meaning and frequency. The information is available in Dutch only.

Leendert Brouwer, editor
Maarten van der Peet et al., programming
Gerrit Bloothoof (Universiteit Utrecht). GBA co-ordination
Jan Pieter Kunst, application components

contact: nfb@meertens.knaw.nl
Corpus of Family Names in the Netherlands

2007
Bloothoofd 19 [kaart] [%kaart]

1947
Bloothoofd 7 [provinciekaart]

network of related names

aal en verspreiding
analyse en verklaring
documentatie
varianten
externe links

VKS 23/9/10

o.a. national bureau for genealogy
Blootoft 23/9/10

2007
Blootoft 19 [kaart] [%kaart]

1947
Blootoft 7 [provinciekaart]

daantal naamdragers per provincie in 1947

- Groningen 0
- Friesland 0
- Drenthe 0
- Overijssel 0
- Gelderland 0
- Utrecht 0
- Amsterdam 0
- Noord-Holland 7
- Noord-Holland totaal 7
- Den Haag 0
- Rotterdam 0
- Zuid-Holland 0
- Zuid-Holland totaal 0
- Zeeland 0
- Noord-Brabant 0
- Limburg 0
- totaal 7
privacy issues

- unique names
  - lesser problem countrywise
  - but unacceptable in small municipalities

- shown: all names
- not shown: any figure < 5
  - Norway: <3, France: <5, Belgium: all

- identifiable:
  - not on map, no colour, for small municipalities (by rules)
  - unless in telephone directory
  - reported to the Commission for the protection of identity data
launching the websites

- Family names: December 3, 2009
  - server failure, > 24 hours
- First names: June 3, 2010
  - no major technical problems

- First day: 3 million page views [google top-10]
- First month: 18 million page views
- High exposure in all media
media

- NOS news (also children’s news)
- all radio channels, newspapers
- hundreds of major websites
- primary school assignments
- …

VKS 23/9/10
why

- everyone has a name
- a name is part of one’s identity
  - how popular is my name
  - where do other people live with my name
  - what is the meaning of my name

- prospective parents:
  - what names have been accepted
  - balanced position with officers of Civil Registration
  - contribution to proposals for liberalisation of name law
we want more…

• geographic distribution per time interval
• migration studies
• extension to 1811
• web 2.0: let everyone contribute
  • origins of the family name
  • etymology of a name
choose your own interval

- Femke (15.205)

1900-1960

1960 - 2009
migration

- we know
  - family relations (in the database of first names)
  - places of birth

- descendants from the oldest generation, where were they born
- ancestors of current generation, where were they born
descendants of people from Sneek

places of birth of children from persons born in Sneek between 1880-1900

This example concerns the town of Sneek, but in the interactive application any municipality can be chosen
descendants of people from Sneek

places of birth of
grandchildren
from persons born in Sneek
between 1880-1900
where do ancestors of people from Sneek come from?

places of birth
from inhabitants of Sneek who are between 30 and 50 years of age
where do ancestors of people from Sneek come from?

places of birth of parents from inhabitants of Sneek who are between 30 and 50 years of age
where do ancestors of people from Sneek come from?

places of birth of grandparents from inhabitants of Sneek who are between 30 and 50 years of age
where do ancestors of people from Sneek come from?

places of birth
of great-grandparents
from inhabitants of Sneek who are between 30 and 50 years of age
let everyone contribute

• enormous amount of distributed knowledge
  • genealogy is among the most popular subjects on the web
  • many people explore archives and possibly know:
    – oldest forms of a family name, circumstances and interpretations
    – development of variants and their spread
    – explicite archival sources

• requires
  – structured information
  – moderation
in conclusion

• names are exciting
• playground for scientific research

challenges for:
• pattern recognition
• space-time representations
• linguistic and socio-onomastic studies
• massive web 2.0 explorations
• (inter)national data sharing (& privacy issues)